

学会企画シンポジウム

医療 AIに向けた医療画像データベース基盤構築の新しい動向

2017年11月21日(火) 14:15 ～ 16:15 E会場 (10F 会議室1003)

[2-E-2-AS1-2] AI等の利活用を見据えた病理組織デジタル画像（WSI）の収集基盤整備と病理支援システム開発ー Japan Pathology AI Diagnostics Project (JP-AID)ー

佐々木 毅¹, 宇於崎 宏², 吉澤 明彦³, 坂元 宇亨⁴, 深山 正久⁵ (1.東京大学医学部附属病院 地域連携推進・遠隔病理診断センター, 2.帝京大学医学部 病理学講座, 3.京都大学医学部附属病院 病理診断科, 4.慶應義塾大学 病理学分野, 5.東京大学大学院医学系研究科 人体病理学・病理診断学分野)

日本病理学会では、日本医療研究開発機構（AMED）の支援により「病理組織デジタル画像（Pathology-Whole Slide Imaging: P-WSI；バーチャルスキャナーで作製した病理デジタル画像）」を、オールジャパン体制でクラウドデータサーバに収集・集約化し、病理診断生涯教育用コンテンツおよび人工知能（Artificial Intelligence：AI）を活用した病理診断支援ツールの開発研究を行う。

現在、病理学会認定医療機関には約100万症例の P-WSIが保存されている。このうち、今年度は大学病院など23医療機関、病理学会7支部等から11万症例の P-WSIおよびそれに紐づいた患者情報等を、病理学会・National Clinical Database（NCD）共用のクラウドデータサーバに収集する。1枚の P-WSIは大容量（9億から36億画素）であるため、収集基盤には学術情報ネットワーク SINET5を使用する。ビッグデータを活用して、病理診断生涯教育用コンテンツ等を NCD、東京大学の支援により共同開発すると同時に、病理学研究支援等を目的にデータ整備、アーカイブ化事業を行う。また AI開発研究用「アノテーション付き正解教師データ」を作製し、国立情報学研究所（National Institute of Informatics: NII）との共同により「AI病理診断支援ツール」の開発および AI研究者へのデータ提供を目的としたアーカイブ化事業も行う。さらに自律性・持続性・悉皆性を備えた収集システムの開発を目指し、「地域基盤・循環型病理診断相互支援モデル」を、滋賀県と長野県の2拠点で実証実験し、P-WSIと患者情報を前向きに収集するためのシステム開発研究も行う。

なお今秋からは、日本消化器内視鏡学会および日本医学放射線学会との画像連携に関する模索研究も計画している。

医療 AI に向けた医療画像データベース基盤構築の新しい動向

大江和彦^{*1}

^{*1} 日本医療情報学会

New Trend of the establishment of medical image database infrastructure for medical AI

Kazuhiko Ohe^{*1}

^{*1} Japan Association of Medical Informatics

Abstract: Deep learning is one of the key technologies for current artificial intelligence and correctly annotated big real world database is another key resource for supervised machine learning. Although development of medical image AI needs such annotated big database, there is no big image database that can be used for AI research so far. This session is organized by JAMI to introduce one of the Japan AMED's project "A research for establishment of medical image database infrastructure toward medical AI". Four invited speakers from academic societies that are involved in this research will introduce each database project, the plan, and the issues to be discussed.

Keywords: Medical AI, medical image database, Deep Learning, National AI infrastructure

1. 医療 AI における画像データベース

多層ニューラルネットワークを用いた機械学習すなわちDeep Neural Networkにより画像識別や画像認識の飛躍的な性能向上がもたらされた。周知のように医療ではとりわけ診断において画像データの重要性が高く、多くの医療分野の画像診断でディープラーニングを活用して診断支援、病変領域の識別、所見分類などが試みられている。こうした研究では教師あり機械学習が用いられており、その教師データすなわち学習データセットとして、正しい診断名や所見名が正解ラベルとして画像に付与された医療画像データが大量に必要である。

このような動きが加速するなかで、平成28年度には日本医療研究開発機構(AMED)の臨床研究等 ICT 基盤構築研究事業において人工知能等利用活用基盤構築関連の研究課題として「ICT 技術や人工知能(AI)等による利活用を見据えた、診療画像等データベース基盤構築に関する研究」が公募された。この公募事業では順不同で(1)日本消化器内視鏡学会「全国消化器内視鏡診療データベースと内視鏡画像融合による新たな統合型データベース構築に関する研究」、(2)日本病理学会「AI 等の利活用を見据えた病理組織デジタル画像(WSI)の収集基盤整備と病理支援システム開発」、(3)九州大学(日本医学放射線学会)「画像診断ナショナルデータベース実現のための開発研究」、の3件が採択された。

従来はこうした採択課題は別々の情報インフラを構築し利用することが多かったと思われるが、今回は国立情報学研究所が分担協力してこの3事業の画像データベースストレージとネットワークの共通インフラを構築することになっていることが大きな特徴である。

日本医療情報学会では学会企画としてこの事業に参加する3団体と国立情報学研究所から演者を推薦していただき、学会企画として本セッションを開催することにした。次項以下では、まず AMED の募集要項から目的や方向性に関する記載を抜粋して紹介した後、3つの各事業の担当発表者と国立情報学研究所の合田氏の抄録を転載することにより、各事業の概要を紹介することにする。

2. テーマの目標と期待される成果

公募要領には「ICT 技術を用いて、悉皆性をもって画像データや診療情報を収集する基盤を構築し、蓄積されたデータから医療の質の向上につながるエビデンスやサービスを創出し、データ収集のインセンティブを確保することに加えて、収集された診療画像等ビッグデータを 2 次利用した人工知能(AI)等の開発にも結びつけること等を目標とします」と書かれている。

また、求められる成果としては「①全国展開を見据えた上で下記に示す体制及び基盤を整備するための方策を研究し、その実践を通じて質の担保された大量の医療画像データ等の収集を行うこと。

- ・医療分野における人工知能(AI)の利活用を促進するため、その利活用が見込まれる医療データを効果的・効率的に収集できる体制、
- ・医療の質の向上や臨床研究の推進にもつながるよう、診療画像データとそれに付随する医療 情報等を、悉皆性を持って自動的に収集するシステム基盤

②収集したデータをもとに人工知能(AI)等の研究開発を行い、将来の普及を見据えつつ、少なくとも特定の分野において実用可能なレベルまで性能を高めた、診療支援ソフトや診療補助 システムのプロトタイプを提案すること。」との記載がある。

本事業がどのような方向性をもっているかを理解する上で、上記の2つの文章は参考になるであろう。

3. 全国消化器内視鏡診療データベースと内視鏡画像融合による新たな統合型データベース構築に関する研究

まず日本消化器内視鏡学会からは田中聖人氏(京都第二赤十字病院)から講演いただくことにし、以下の抄録をいただいた。「消化器内視鏡機器のシェアは世界の 70%を日本製品が占めており、学術的な研究、技術の先進性はわが国が世界をリードしている。国内の年間の消化器内視鏡診療件数は約 1300 万件で、診療画像は億を超える枚数であるが、臨床研究の基礎となる膨大な診療データと画像を集積する大規模データベースが存在しない。また、内視鏡診療においては、画像診断という知的習熟が必須で、人工知能による画像診

断等の支援システムによる、術者を選ばない正確な診療の均霑化が求められている。このような現状を踏まえ、消化器内視鏡診療領域においても全国統計・臨床研究のレジストリとしての悉皆性を持ったデータ収集が可能な情報基盤の整備と、膨大な診療情報・画像を統合管理するデータベースの構築が急がれる。今回の研究では診療情報と画像を各ベンダーが提供する内視鏡部門システムへ保存するという内視鏡診療の特性を生かし、大規模なデータベース構築の基盤を形成する。

しかし、ベンダー側でデータ外部出力やデータ構造・用語に関する標準化がまったく考慮されておらず、部門システムを介してデータを収集する際の大きな障壁となっていた。そこで、各診療施設の内視鏡部門システムから、症例データと画像データの抽出後に個人情報自動で除去し、学会サーバへ送信する施設側ゲートウェイアップローダを開発し、部門システムから直接データを抽出加工することにより、二重入力無しにデータの信頼性が担保され、データ収集プロセスの煩雑な作業を可能な限り自動化する事を目的としている。

こうして得られたものによって①医療の質向上、安全に関するデータ分析研究、②集積された大規模症例データの解析により臨床研究の次元が拡がり、③疾患予防に繋がる背景情報の検証や、④偶発症リスクの低減が可能となる。

本シンポジウムでは現在の進捗状況と問題点、そして今後長期的に広がる未来に関して概説する。」

4. AI等の利活用を見据えた病理組織デジタル画像(WSI)の収集基盤整備と病理支援システム開発

日本病理学会からは佐々木毅氏(東京大学医学部附属病院)に講演いただくこととし、以下の抄録をいただいた。「日本病理学会では、日本医療研究開発機構(AMED)の支援により「病理組織デジタル画像(Pathology-Whole Slide Imaging: P-WSI);バーチャルスキャナーで作製した病理デジタル画像)」を、オールジャパン体制でクラウドデータサーバに収集・集約化し、病理診断生涯教育用コンテンツおよび人工知能(Artificial Intelligence: AI)を活用した病理診断支援ツールの開発研究を行う。

現在、病理学会認定医療機関には約100万症例のP-WSIが保存されている。このうち、今年度は大学病院など23医療機関、病理学会7支部等から11万症例のP-WSIおよびそれに紐づいた患者情報等を、病理学会・National Clinical Database(NCD)共用のクラウドデータサーバに収集する。1枚のP-WSIは大容量(9億から36億画素)であるため、収集基盤には学術情報ネットワークSINET5を使用する。ビッグデータを活用して、病理診断生涯教育用コンテンツ等をNCD、東京大学の支援により共同開発すると同時に、病理学研究支援等を目的にデータ整備、アーカイブ化事業を行う。またAI開発研究用「アノテーション付き正解教師データ」を作製し、国立情報学研究所(National Institute of Informatics: NII)との共同により「AI病理診断支援ツール」の開発およびAI研究者へのデータ提供を目的としたアーカイブ化事業も行う。さらに自律性・持続性・悉皆性を備えた収集システムの開発を目指し、「地域基盤・循環型病理診断相互支援モデル」を、滋賀県と長野県の2拠点で実証実験し、P-WSIと患者情報を前向きに収集するためのシステム開発研究も行う。

なお今秋からは、日本消化器内視鏡学会および日本医学放射線学会との画像連携に関しての模索研究も計画してい

る。」

5. 画像診断ナショナルデータベース(Japan Medical Imaging database: J-MID)について

日本医学放射線学会からは待鳥詔洋氏(国立国際医療研究センター国府台病院)に講演いただくこととし、以下の抄録をいただいた。「画像診断技術の普及や進歩は、医療の高度化、患者の予後改善に多大な貢献をしてきた。その技術革新の進歩は非常に早く、今後はICT技術を取り入れ新たなステージに進むことが期待される。我が国の医療は、少子高齢化、国の財政悪化等の問題に直面しており、画像診断のビッグデータを利活用し、医療費増大に対応しつつ医療技術の向上を継続することは、今後の画像診断の発展にとって、避けて通れない課題である。日本医学放射線学会は、「Japan Safe Radiology」を提唱した。これは、今回構築される画像診断ナショナルデータベース(Japan Medical Imaging database: J-MID)を解析・活用し、医療安全の向上、標準化、適正化、人工知能(AI)等を含む新たな技術開発、医療政策への反映などを狙った構想である。日本医学放射線学会が実施するナショナルデータベース事業について概説する。」

6. 医療画像ビッグデータクラウド基盤

最後の演者であり、共通インフラを担当する国立情報学研究所からは合田憲人氏(国立情報学研究所)に講演いただくこととし、以下の抄録をいただいた。「人工知能(AI)技術を用いた画像解析技術の向上に伴い、医療画像の解析にAIを活用することへの期待が高まっている。本講演では、国立情報学研究所が中心となって構築を進めている医療画像ビッグデータクラウド基盤について紹介する。本クラウド基盤は、AI技術を活用した医療画像ビッグデータ解析研究のための計算基盤であり、研究者は、医療系学会等の拠点から送付される匿名化された医療画像ビッグデータをAI技術を活用して解析することが可能である。医療系学会等の拠点と本クラウド基盤とを接続するネットワークには、学術情報ネットワークSINET5を用いる。SINET5は、全国850以上の大学や研究機関等が接続された学術専用的高性能ネットワークであり、全都道府県を100Gbpsで接続する高速性や、インターネットと切り離されたVirtual Private Network(VPN)を柔軟に構成できる等の高機能性を備えている。Deep Learning等の手法を用いた最新の画像解析技術では非常に高い計算能力が必要とされるが、本クラウド基盤では、高速なDeep Learning処理を可能とする高性能GPUを搭載したサーバを利用することにより、研究者がAI画像解析を高速に実施することが可能である。」

7. 考察

本事業によりデータベース基盤の構築はまだ始まったばかりであるから本学会としては各データベース事業の構想や規模感と将来計画などを議論していきたい。そのなかで議論のポイントとしては(1)悉皆性を高めていくためにはどのような体制が必要か、(2)研究事業後の継続性をどのように維持していくか、(3)アノテーションラベルの課題、すなわち個々の画像に対する所見情報(正解ラベル)をどのように付与していくか、またその用語やコードの標準化、情報の細かさや部位に関する情報の粒度の共通化などをどのように決定し、継続的な付与作業をどのようにおこなっていくのか、(4)まずはどのようなAI応用の実現を目指すか、などが考えられよう。